arXiv:math/0311226v2 [math.NT] 26 Nov 2003

A Combinatorial Method for Counting Smooth Numbers in Sets of Integers

Ernie Croot

February 1, 2008

Abstract

In this paper we prove a result for determining the number of integers without large prime factors lying in a given set S. We will apply it to give an easy proof that certain sufficiently dense sets A and B always produce the expected number of "smooth" sums $a + b$, $a \in A$, $b \in B$. The proof of this result is completely combinatorial and elementary.

1 Introduction

Given a set S, a common question one tries to answer is whether S contains the expected number of "y-smooth" integers, which are those integers having no prime divisors greater than y. We denote the number of integers in S with this property by $\Psi(S, y)$; and for a number $x > 0$, one denotes the set of all y-smooths positive integers $\leq x$ by $\Psi(x, y)$. So,

$$\Psi(\{1, 2, ..., \lfloor x \rfloor\}, y) \ = \ \Psi(x, y).$$

If $S \subseteq \{1, 2, ..., x\}$, then, all things being equal, one would expect that

$$\frac{\Psi(S, y)}{|S|} \ \sim \ \frac{\Psi(x, y)}{x}. \tag{1}$$

For example, fix a real number $0 < \theta \leq 1$ and an integer $a \neq 0$, and let S be the set of numbers of the form $p + a$, where $p \leq x$ runs through the primes; S is often called a set of "shifted primes". It is conjectured that

$$\Psi(S, x^\theta) \ \sim \ \frac{\pi(x)\Psi(x, x^\theta)}{x} \ \sim \ \rho(\theta^{-1})\pi(x), \tag{2}$$

where

$$\rho(u) \;=\; \lim_{x\to\infty} \frac{\Psi(x, x^{1/u})}{x}.$$

This function ρ is called Dickman's function, and it was proved in [7] that the limit exists. Unfortunately, proving (2) remains a difficult, open problem; however, in [9], J. B. Friedlander gave a beautiful proof that $\Psi(S, x^\theta) \gg \pi(x)$ for $\theta > (2\sqrt{e})^{-1}$, and in [1], R. Baker and G. Harman proved that for $\theta \geq 0.2961$,

$$\Psi(S, x^\theta) \;>\; \frac{x}{\log^\alpha x},$$

for some $\alpha > 1$ and $x > x_0(a)$.

There are several methods for attacking the general question of proving that (1) holds for a particular set S, one such method involves exponential sums and the circle method, and another uses a Buchstab identity, in combination with a sieve method (such as the Large Sieve).

In this paper we offer a novel way of showing that sets S have the expected number of x^θ-smooths, and the conditions that S needs to satisfy, in order for this method to work, are simpler than those required by other methods (such as an application of Buchstab's identity). Before we can state what these conditions are, we first introduce the notion of a "Local-Global Set", which we abbreviate as LG set:

Definition. We say that $N \subseteq \{2, 3, ..., x\}$ is an LG set with parameters ϵ, c and x if and only if following two conditions hold:

1. For any pair of distinct members $n_1, n_2 \in N$ we have $\mathrm{lcm}(n_1, n_2) > x$;
2. All but at most ϵx of the integers $m \leq x$ are divisible by some $n \leq x^c$ with $n \in N$.

Notes: From condition 1 we know that if $m \leq x$ is divisible by one of these n's, then this n must be unique, else if $n_1, n_2 \in N$ are distinct and if $n_1|m$ and $n_2|m$, then $\mathrm{lcm}(n_1, n_2)|n$, which implies $n \geq \mathrm{lcm}(n_1, n_2) > x$, contradiction. We also note that condition 2 implies

$$\begin{aligned}
\sum_{\substack{n\in N\\ n<x^c}} \frac{1}{n} \quad &= \quad \frac{1}{x}\sum_{\substack{n\in N\\ n<x^c}} \frac{x}{n} \\
&= \quad \frac{1}{x}\sum_{\substack{n\in N\\ n<x^c}} \left(\#\{m \leq x \;:\; n|m\} + O(1)\right)
\end{aligned}$$

$$\begin{aligned} &= \frac{1}{x}\#\{m \le x \ :\ \exists n \in N,\ n < x^c,\ \text{where } n|m\} \ +\ O(x^{c-1}) \\ &= 1 - \epsilon', \end{aligned} \tag{3}$$

where $0 < \epsilon' < 2\epsilon$ for x sufficiently large.

The main result we will use in the development of our "smooth sieve method" is the following theorem:

Theorem 1 *For every $0 < \epsilon < \epsilon_0$ (for some ϵ_0) and for x sufficiently large (in terms of ϵ) there exists an LG set of integers $N \subseteq \{1, 2, ..., x\}$ with parameters ϵ, $c = c(\epsilon)$ (that is, c depends only on ϵ, and not on x) and x. Moreover, the following is an explicit example of such a set: For a certain constant δ, depending only on ϵ, and for x sufficiently large, we let N be the set of integers n of the form $p_1 p_2 \cdots p_k \le x$ (k variable), where the p_i's are prime numbers such that*

$$p_1 > p_2 > \cdots > p_k > x^\delta,$$

and

$$\text{For i} = 1, 2, ..., \text{k} - 1,\ \frac{x}{p_1 p_2 \cdots p_i} \ge p_i; \quad \text{and } 1 \le \frac{x}{p_1 p_2 \cdots p_k} < p_k. \tag{4}$$

Remark 1: For the explicit construction given, in order for condition 2 for being an LG set to be satisfied, we need only choose δ so small that

$$\sum_{n \in N} \frac{1}{n} \ > \ 1 - \frac{\epsilon}{2}.$$

To see this, we note that since each member of N has all its prime divisors $> x^\delta$, basic sieve methods show that for $y > x^{1/2}$,

$$\sum_{\substack{n \in N \\ n \in [y, 2y]}} \frac{1}{n} \ < \ \frac{C_\delta}{\log x},$$

where C_δ depends only on δ. Summing over dyadic intervals, this implies

$$\sum_{\substack{n \in N \\ x^c < n < x}} \frac{1}{n} \ < \ D_\delta(1 - c),$$

for a certain constant D_δ depending only on δ. Now, by taking c sufficiently close to 1, this sum can be made less than $\epsilon/2$, which would give

$$\sum_{\substack{n\in N\\ n<x^c}} \frac{1}{n} \ > \ 1-\frac{\epsilon}{2}-\sum_{\substack{n\in N\\ x^c\leq n<x}} \frac{1}{n} \ > \ 1-\epsilon.$$

The main theorem of the paper is the following result:

Theorem 2 *Given $0<\theta\leq 1$ and $\gamma>0$, there exists $0<\epsilon<1$ so that for x sufficiently large, if N is an LG set with parameters ϵ, c, and x as given in Theorem 1, then the following holds: First, let N_1 be the set of x^θ-smooths that lie in N and are $<x^c$; let N_2 be the set of integers that are not x^θ-smooth that lie in N and are $<x^c$; let $w(n)\geq 0$ be some weighting function on positive integers $\leq x$; and let*

$$\sigma \ = \ \sum_{s\leq x} w(s).$$

Further, suppose that the following two inequalities hold

$$\sum_{q\in N_1}\sum_{\substack{s\leq x\\ q|s}} w(s) \ > \ (1-\gamma)\sigma\sum_{q\in N_1}\frac{1}{q}; \text{ and} \tag{5}$$

$$\sum_{q\in N_2}\sum_{\substack{s\leq x\\ q|s}} w(s) \ > \ (1-\gamma)\sigma\sum_{q\in N_2}\frac{1}{q}. \tag{6}$$

Then, we have that

$$\left|\sum_{\substack{s\leq x\\ s \text{ is } x^\theta-\text{smooth}}} w(s) \ - \ \sigma\sum_{q\in N_1}\frac{1}{q}\right| \ < \ 2\gamma\sigma.$$

Moreover, if one is only able to show (5), then one can deduce the sharp lower bound

$$\begin{aligned}\sum_{\substack{s\leq x\\ s \text{ is } x^\theta-\text{smooth}}} w(s) \quad &> \quad (1-\gamma)\sigma\sum_{q\in N_1}\frac{1}{q}\\ &> \quad (1-\gamma)(\rho(1/\theta)-2\epsilon)\sigma,\end{aligned}$$

for sufficiently large x.

Remark. For a fixed $\epsilon > 0$, as x tends to infinity the sum of $1/q$ over $q \in N_1$ tends to a limit which is within 2ϵ of $\rho(1/\theta)$. So, as ϵ tends to 0 and x tends to infinity, the sum of $1/q$ over $q \in N_1$ tends to the limit $\rho(\theta^{-1})$, where ρ is Dickman's function, and was stated earlier.

Let us now discuss this theorem, to get a feel for how one might use it. What this theorem is saying is that if one has a set of integers S, and if one can produce only good lower bounds (upper bounds are not needed) for how many elements in S there are that are divisible by elements $q \in N_1$, and by elements $q \in N_2$, then one can deduce that S contains the expected number of x^θ-smooths. One of the strengths of the theorem is that the q's for which one needs the divisibility conditions (5) and (6) to hold are less than a power of x (where the power is less than 1). With a naive Buchstab identity approach one *must* confront the problem of whether there are lots of elements in the set S that are divisible by large primes; and, even if one can solve that problem, there are still other problems involving divisibility by large q's that must be addressed. These difficulties never need to be confronted in the theorem above.

Another strength is that we only require lower bounds on the left-hand-side sums in (5) and (6), and it is sometimes much easier to produce such lower bounds, than it is to give asymptotic estimates. A fairly simple example is the following: Suppose that one takes $\theta \in (0, 1]$ and $\gamma > 0$ "close" to 0, and lets $c = c(\theta, \gamma)$ be as in the theorem above. Now take A to be a subset of the integers $\leq x$ having at least $x^{c+\nu}$ elements, where $0 < \nu < 1$ can be taken arbitrarily small. Let $w(n)$ be the number of ways of writing n as a difference of two elements of A. The sum of $w(n)$ over all positive integers n is obviously $\binom{|A|}{2}$. It is also easy to see that the sum over all $w(n)$ with $n \geq 1$ divisible by q is

$$\sum_{a=0}^{q-1} \binom{A(a,q)}{2},$$

where $A(a, q)$ is the number of elements of A that are $\equiv a \pmod{q}$. This expression is minimized if the elements of A are as equidistributed amongst the residue classes modulo q as is possible; and so, this expression can be shown to be at least $\sim |A|^2/(2q)$ in size. Summing over all n divisible by

elements $q \in N_i$ of $w(n)$, we get that

$$\sum_{q \in N_i} \sum_{\substack{n \geq 1 \\ q | n}} w(n) \ \gtrsim \ \frac{|A|^2}{2} \sum_{q \in N_i} \frac{1}{q}.$$

Thus, (5) and (6) hold for x sufficiently large, and it follows that the number of differences $a - b > 0$, $a, b \in A$, which are x^θ-smooth is "close" to $U = (|A|^2/2) \sum_{q \in N_1} 1/q \approx \rho(1/\theta)|A|^2/2$; and, the smaller we take γ to be, the closer this count will be to U. In the next section, we will give a proof of a related (but more difficult) result.

The proof of theorem 2 is so simple that we will give it here in the introduction:

Proof of Theorem 2. First, we want the value of ϵ for the LG set to be so small that $\theta > \delta$, where δ is the parameter given in the construction in Theorem 1. There will be additional demands on ϵ that we will give as the proof progresses.

Now suppose $n \leq x$ is divisible by some $q \in N$ (which is unique). Then, we have that n is x^θ-smooth if and only if $q \in N_1$. To see this, first note that if $n = qk$, where $q \in N_1$, then by the definition of the set N, we have that x/q is less than the smallest prime dividing q; that is, k is less than x^θ (which must be greater than x^δ). So, n is x^θ-smooth. Conversely, suppose that $n \leq x$ is x^θ-smooth and $q|n$. Then, if q is not x^θ-smooth, then neither is n.

We deduce that the sum of $w(s)$ over all the elements $s \leq x$ that are x^θ-smooths is *at least* the sum of $w(s)$ over all $s \leq x$ divisible by some $q \in N_1$. This quantity is given by the left hand side of (5). On the other hand, the sum of $w(s)$ over all $s \leq x$ that are x^θ-smooth is *at most* $\sigma - \tau$, where τ is the sum of $w(s)$ over all $s \leq x$ divisible by some $q \in N_2$. From (6), we get that this quantity is at most

$$\begin{aligned}
\sigma \ - \ \tau \quad &< \quad \sigma - (1-\gamma)\sigma \sum_{q \in N_2} \frac{1}{q} \\
&< \quad \sigma - (1-\gamma)\sigma \left((1-\epsilon') - \sum_{q \in N_1} \frac{1}{q} \right) \\
&< \quad \sigma \sum_{q \in N_1} \frac{1}{q} \ + \ (\gamma + \epsilon')\sigma,
\end{aligned}$$

where ϵ' is as given in (3), and therefore depends on the value of ϵ (but tends to 0 as ϵ tends to 0). Now, if ϵ is sufficiently small, then ϵ' will be smaller than γ, and so this last chain of inequalities would give

$$\sigma \ - \ \tau \ < \ \sigma \sum_{q \in N_1} \frac{1}{q} \ + \ 2\gamma\sigma.$$

Combining this with (5) then proves the theorem. ■

The remainder of this paper is organized as follows. In the next section we give an application of Theorem 2 to counting the number of smooth sums $a + b$, where a lies in some set of integers A, and b lies in a set B, and in section 3, we will give a proof of Theorem 1.

2 An Application of Theorems 2

Given sets of integers A and B, which are subsets of $\{1, 2, ..., x\}$ having $\gg x$ elements each, it is an interesting and studied question to determine the number of y-smooth sums $a + b$, $a \in A$, $b \in B$. There are several ways of attacking this sort of problem, one of which is to use the circle method and exponential sums over smooth numbers, and another is to use the large sieve. We could also ask how $\tau(a + b)$ is distributed, or how large $P(a + b)$ can be, where τ is the divisor function, and where $P(n)$ denotes the largest prime factor of n. Using the large sieve and the circle method, these types of questions were given a thorough treatment in a series of beautiful papers by A. Balog and A. Sarkozy [2], [3], [4], and [5]; P. Erdős, H. Maier, and A. Sárkőzy [8]; A. Sárkőzy and C. L. Stewart [11], [12], [13], [14]; C. Pomerance, A. Sárkőzy, and C. L. Stewart [10]; and R. de la Bretèche [6]. The paper by de la Brèteche is more relevant to the main result of this section, and we give here one of his theorems:

Theorem 3 *Suppose that A and B are subsets of the integers in $\{1, 2, ..., x\}$. For a given integer $y \le x$, let $u = (\log x)/\log y$. Then, uniformly for $x \ge 3$, $\exp((\log x)^{2/3+\epsilon}) < y \le x$ we have*

$$\begin{aligned} &\#\{a \in A,\ b \in B\ :\ P(a+b) \le y\} \\ &\qquad = \ |A| \cdot |B| \rho(u) \left(1 + O\left(\frac{x}{\sqrt{|A| \cdot |B|}} \frac{\log(u+1)}{\log y}\right)\right), \end{aligned}$$

where $P(n)$ denotes the largest prime divisor of n, and where ρ is Dickman's function.

R. de la Bretèche used estimates for exponential sums and the circle method to prove this result. Notice that if $|A| \cdot |B| \ll (x/\log x)^2$, then his result fails to prove that there are the expected number of sums that are y-smooth for any $y < x$, because in this case the big-Oh term is $\gg 1$.

Let us now consider what happens in the case when $y = x^\theta$ (and so $u = 1/\theta$): Is it possible to show that if $|A|, |B| > x^c$, for some $0 < c < 1$, then we get the expected number of sums $a + b$ being y-smooth? It is easy to see that the answer is no, no matter how close to 1 we take c to be. For example, we could take A and B to both be the set of integers $\leq x$ that are divisible by some prime number p around size x^{1-c}. Notice here that $|A| \sim x^c$. But then, the sums $a + b$, $a, b \in A$ are numbers of the form pk, where $k < 2x^c$, and such a sum is x^θ-smooth if and only if k is x^θ-smooth, when $p < x^\theta$. Thus, one would expect (and can show) that

$$\frac{\#\{a, b \in A \ : \ a + b \text{ is } x^\theta - \text{smooth}\}}{|A| \cdot |B|} \ \sim \ \frac{\Psi(2x^c, x^\theta)}{2x^c} \ \sim \ \rho(c/\theta).$$

On the other hand, the proportion of x^θ smooths $\leq x$ is $\sim \rho(u')$, where $u' = \log(x)/\log(x^\theta) = 1/\theta$, which is not c/θ. So, the type of result we might try to prove is the following:

Theorem 4 *Given $0 < \theta \leq 1$, and $\gamma > 0$, there exists $\nu = \nu(\theta, \gamma) \in (0, 1)$ so that the following holds: For x sufficiently large, if $A, B \subseteq \{1, 2, ..., x\}$ satisfy $|A|, |B| > x^\nu$, then*

$$\begin{aligned} \big|\#\{a, b \ : \ a \in A, \ b \in B, \ a + b \ \text{ is } x^\theta - \text{smooth}\} - \rho(1/\theta)|A| \cdot |B|\big| \\ < \ \gamma |A| \cdot |B|. \end{aligned}$$

The same result holds for differences $a - b$ (with possibly a different value for ν).

There are other methods for proving this type of theorem, besides the exponential sums approach used by de la Bretèche, such as an application of Buchstab's identity, together with a form of the large sieve for sieving by composite moduli; however, these methods are technical, and it does not seem possible to give an easy and elegant proof of Theorem 4 using them.

In the remainder of this section we will give an entirely elementary proof of Theorem 4 using Theorem 2. First, though we need a "large sieve"-type result for composite moduli, and although such a result can be easily proved by simply modifying the standard large sieve, we give here (perhaps astonishingly) a completely elementary proof based on LG sets.

Theorem 5 *Given $\epsilon > 0$ and x sufficiently large, let N be an LG set for parameters ϵ and x, as given in Theorem 1. Further, suppose that $c = c(\epsilon)$ is as in property 2 for being an LG set. Suppose that $C \subseteq \{1, 2, ..., x\}$, and let $C(a, q)$ denote the number of elements of C that are congruent to a modulo q. Then, we have that*

$$\sum_{\substack{q \in N \\ q < x^c}} \sum_{a=0}^{q-1} \left(C(a, q) - \frac{|C|}{q} \right)^2 \ < \ |C|(2\epsilon|C| + x^c).$$

The proof of this "Large Sieve"-like theorem is so simple, we will not postpone its proof to a later section:

Proof. To make the notation simple, when we sum over $q \in N$, we mean the sum over those $q \in N$ satisfying $q \leq x^c$.

We note that if $b, c \in C$, and $b \neq c$, then if $q \in N$ divides $b - c$, we must have that q is unique; otherwise, if $q' \in N$ also divides $b - c$, then $\mathrm{lcm}(q, q') > x$ divides $b - c$, which is impossible.

Thus, we have that

$$\begin{aligned} |C|^2 \quad &> \quad \sum_{q \in N} \#\{b, c \in C, \ b \neq c \ : \ q|(b - c)\} \\ &= \quad \sum_{q \in N} \sum_{a=0}^{q-1} \left(C(a, q)^2 - C(a, q) \right) \\ &= \quad \sum_{q \in N} \sum_{a=0}^{q-1} C(a, q)^2 \ - \ x^c|C|. \end{aligned}$$

Thus,

$$|C|(x^c + |C|) \ > \ \sum_{q \in N} \sum_{a=0}^{q-1} C(a, q)^2.$$

It now follows that

$$\begin{aligned}\sum_{q\in N}\sum_{a=0}^{q-1}\left(C(a,q)-\frac{|C|}{q}\right)^2 &= \sum_{q\in N}\sum_{a=0}^{q-1}C(a,q)^2-|C|^2\sum_{q\in N}\frac{1}{q}\\ &< (1-(1-\epsilon'))|C|^2+x^c|C|\\ &< |C|(2\epsilon|C|+x^c),\end{aligned}$$

where ϵ' is as in (3). The theorem is now proved. ∎

Proof of Theorem 4. Given $0<\theta\le 1$, and $\gamma>0$, we suppose ϵ is so small that the conclusion of Theorem 2 holds for x sufficiently large. Let N be the LG set with parameters ϵ, $c=c(\epsilon)$ and x (for x sufficiently large) as appears in the construction in Theorem 1. Finally, let $\delta=\delta(\epsilon)$ be the parameter also given in this construction, and let N_1 and N_2 be the sets as described in the conclusion of Theorem 2. Two additional demands on ϵ and x is that we will need ϵ to be so small, and x so large that

$$\left|\sum_{n\in N_1}\frac{1}{n}\ -\ \rho(1/\theta)\right|\ <\ \frac{\gamma}{4}, \tag{7}$$

and

$$\epsilon\ <\ \frac{\gamma}{12}\sum_{q\in M}\frac{1}{q},\ \text{ for } M=N_1 \text{ and } N_2.$$

We will show that the conclusion of our theorem holds for any $\nu>c=c(\epsilon)$ for x sufficiently large.

Let α be the indicator function on the set A, let β be the indicator function on the set B, let

$$A(a,q)\ =\ \sum_{\substack{n\le x\\ n\equiv a \pmod q}}\alpha(n),\ \text{ and } B(a,q)\ =\ \sum_{\substack{n\le x\\ n\equiv a \pmod q}}\beta(n),$$

and, finally, define the weight function

$$w(n)\ =\ \sum_{a+b=n}\alpha(a)\beta(b).$$

Then, for $M = N_1$ or N_2, we get

$$\begin{aligned}\sum_{q\in M}\sum_{\substack{n\le x\\ q|n}} w(n) &= \sum_{q\in M}\sum_{a=0}^{q-1} A(a,q)B(q-a,q)\\ &= \sum_{q\in M}\sum_{a=0}^{q-1}\left(A(a,q)-\frac{|A|}{q}\right)\left(B(q-a,q)-\frac{|B|}{q}\right)\\ &\qquad + |A|\cdot|B|\sum_{q\in M}\frac{1}{q}. \qquad (8)\end{aligned}$$

Now, then, to bound the last double sum in (8) from above we apply the Cauchy-Schwarz inequality, together with Theorem 5 with $C = A$ and $C = B$:

$$\begin{aligned}&\sum_{q\in M}\sum_{a=0}^{q-1}\left|A(a,q)-\frac{|A|}{q}\right|\ \left|B(q-a,q)-\frac{|B|}{q}\right|\\ &\le \left(\sum_{q\in M}\sum_{a=0}^{q-1}\left(A(a,q)-\frac{|A|}{q}\right)^2\right)^{1/2}\left(\sum_{q\in M}\sum_{a=0}^{q-1}\left(B(a,q)-\frac{|B|}{q}\right)^2\right)^{1/2}\\ &\le (|A|(2\epsilon|A|+x^c))^{1/2}(|B|(2\epsilon|B|+x^c)^{1/2}\\ &\le 3\epsilon|A|\cdot|B|, \qquad (9)\end{aligned}$$

for $|A|, |B| > x^c/\epsilon$, which certainly holds if $\nu > c$ and x is sufficiently large.

Combining this with (8) we deduce that

$$\begin{aligned}\sum_{q\in M}\sum_{\substack{n\le x\\ q|n}} w(n) &> |A|\cdot|B|\left(-3\epsilon+\sum_{q\in M}\frac{1}{q}\right)\\ &> |A|\cdot|B|\left(1-\frac{\gamma}{4}\right)\sum_{q\in M}\frac{1}{q}. \qquad (10)\end{aligned}$$

for $M = N_1$ or N_2. Thus, the conditions of Theorem 2 are met, and we deduce that if σ is the sum of $w(n)$ over all $n \le x$, which is $|A|\cdot|B|$, then

$$\begin{aligned}&|\#\{a,b\ :\ a\in A,\ b\in B,\ a+b \text{ is } x^\theta-\text{smooth}\}-\rho(1/\theta)\sigma|\\ &\le \left|\#\{a,b\ :\ a\in A,\ b\in B,\ a+b \text{ is } x^\theta-\text{smooth}\}-\sigma\sum_{q\in N_1}\frac{1}{q}\right|+\frac{\gamma}{4}\sigma\end{aligned}$$

$$
\begin{aligned}
&< \frac{\gamma}{2}\sigma + \frac{\gamma}{4}\sigma \\
&< \gamma\sigma.
\end{aligned}
$$

The theorem now follows. ■

3 Proof of Theorem 1

First, we show that the set N described in the statement of the theorem satisfies the first condition for being an LG set, namely that for any distinct pair of integers $n_1, n_2 \in N$, we have $\mathrm{lcm}(n_1, n_2) > x$: Given such n_1, n_2, write out their prime factorizations as

$$
\begin{aligned}
n_1 &= p_1 \cdots p_k,\ p_1 > p_2 > \cdots > p_k;\ \text{and} \\
n_2 &= q_1 \cdots q_\ell,\ q_1 > q_2 > \cdots > q_\ell.
\end{aligned}
$$

Without loss of generality, we can assume that $p_k \leq q_\ell$.

Now, if there is some prime q_i which is distinct from the primes $p_1, ..., p_k$, then we would have that the lcm of n_1 and n_2 is divisible by the product of primes $q_i p_1 \cdots p_k$, and this product exceeds x, because from (4)

$$
q_i p_1 \cdots p_k \ >\ q_i \frac{x}{p_k}\ \geq\ q_\ell \frac{x}{p_k}\ \geq\ x.
$$

So we are left to consider what happens when the q_i's are a subset of the p_i's. We break this case into two sub-cases, with the first one where $p_k = q_\ell$, and the second where $p_k < q_\ell$.

In the case $p_k = q_\ell$, we must have that there exists one of the primes $p_i > q_\ell$ such that p_i is distinct from $q_1, ..., q_\ell$, since otherwise we would have $n_1 = n_2$. But now our assumption gives

$$
\frac{x}{p_1 \cdots p_k}\ \leq\ \frac{x}{p_i q_1 \cdots q_\ell}\ <\ \frac{q_\ell}{p_i}\ <\ 1,
$$

which is impossible.

So, we may assume $p_k < q_\ell$. For this case, let $j < k$ be the index where $p_j = q_\ell$ (which exists since q_i's are a subset of the p_i's). Then, we have

$$
p_1 \cdots p_j\ \geq\ q_1 \cdots q_\ell.
$$

From (4) this gives

$$q_\ell \;=\; p_j \;\leq\; \frac{x}{p_1\cdots p_j} \;\leq\; \frac{x}{q_1\cdots q_\ell} \;<\; q_\ell,$$

which is impossible. So, we conclude that the set described in the statement of the theorem satisfies the first condition for being an LG set.

We are left to show that for some $0<\delta<1$, and all x sufficiently large, all but ϵx of the integers $m\leq x$ are divisible by some member of N. We will do this by identifying a subset $T\subseteq\{1,2,...,x\}$ having at least $(1-\epsilon/2)x$ elements, such that all but at most $\epsilon x/2$ of the elements of T are divisible by some member of N. The way we will show this is to construct a weighting function $f(t)>0$ such that if $t\in T$ is not divisible by any $n\in N$, then $f(t)$ will be "large". But then, we will show that average value of $f(t)$ over all $t\in T$ is "much smaller" these large values; so, it will follow that there can be few integers $t\in T$ not divisible by any $n\in N$.

Let $k=\lfloor 1/\epsilon\rfloor+1$; let $0<\gamma<1$ be some constant to be chosen later; let $H_1,...,H_k$ be intervals given by $H_j=[x^{\gamma^{j+1}},x^{\gamma^j})$; let $I_1,...,I_k$ be the intervals where $I_j=[x^{\gamma^{j+1}},x^{\gamma^j/2}]$; and finally, let $J_1,...,J_k$ be the intervals $J_j=(x^{\gamma^j/2},x^{\gamma^j})$. Note here that $H_j=I_j\cup J_j$. Our constant δ in the construction in the statement of the main theorem will be $\delta=\gamma^{k+1}$.

We first claim that for x sufficiently large, all but at most $\epsilon x/4$ integers $m\leq x$ satisfy the following inequality for all $j=1,2,...,k$:

$$\sum_{\substack{p^a|m,\ p \text{ prime}\\ p<x^{\gamma^j}}}\log p \;<\; 5k^2\gamma^j\log x, \tag{11}$$

To see this, we first note that for any $j=1,2,...,k$,

$$\begin{aligned}
\sum_{m\leq x}\sum_{\substack{p^a|m,\ p \text{ prime}\\ p<x^{\gamma^j}}}\log p \quad &= \quad \sum_{\substack{p\leq x^{\gamma^j}\\ p \text{ prime}}}(\log p)\sum_{a\geq 1}\#\{m\leq x \;:\; p^a|m\}\\
&\leq \quad \sum_{\substack{p\leq x^{\gamma^j}\\ p \text{ prime}}}(\log p)\sum_{a\geq 1}\frac{x}{p^a}\\
&= \quad x\sum_{\substack{p\leq x^{\gamma^j}\\ p \text{ prime}}}\frac{\log p}{p} \;+\; O(x)\\
&= \quad \gamma^j x\log x \;+\; O(x), \qquad (12)
\end{aligned}$$

where the constant in the last big-oh depends on γ. The last line here was gotten by using the fact that the sum over primes $p \leq x$ of $(\log p)/p$ is $\log x + O(1)$. Now, for x sufficiently large, there can be at most $x/(4k^2)$ integers $m \leq x$ which fail to satisfy (11) for j, since otherwise these exceptional integers n would force the first sum in (12) to be of size at least $5\gamma^j x(\log x)/4$, and we know this sum is of size at most $\gamma^j x(\log x)$. So, for x sufficiently large, (11) holds for all $j = 1, 2, ..., k$ for all but at most

$$k\frac{x}{4k^2} \ = \ \frac{x}{4k} \ < \ \frac{\epsilon x}{4}$$

exceptional integers $m \leq x$. Let S denote the set of integers satisfying (11) for all $j = 1, 2, ..., k$. Then, we have shown $|S| > (1 - \epsilon/4)x$ for x sufficiently large.

Let $h(s)$ be the number of integers $j = 1, 2, ..., k$ such that s is divisible by some prime $p \in J_j$. So, $0 \leq h(s) \leq k$. We will show below that all but at most $\epsilon x/4$ integers $s \leq x$ satisfy

$$\left| h(s) - \frac{k}{2} \right| < k^{2/3};\ s > \frac{\epsilon x}{100};\ \text{and},\ p^2 | s,\ p \text{ prime } \ \Rightarrow\ p < x^{\delta}. \tag{13}$$

These last two conditions are obviously satisfied for all but at most $\epsilon x/50$ integers $s \leq x$ for large x; and so, we just need to show that the first condition holds for all but at most $\epsilon x/5$ integers $n \leq x$. We will then let T be the set of all $s \in S$ satisfying these conditions. Clearly we will have $|T| > (1 - \epsilon/2)x$.

To prove that the first condition of (13) holds for all but at most $\epsilon x/5$ integers $s \leq x$, we begin by supposing $V \subseteq \{1, 2, ..., k\}$ and $W = \{1, 2, ..., k\} \setminus V$, and letting U denote the set of all integers $u \leq x$ such that

1. For every $v \in V$, u is not divisible by any prime $p \in J_v$; and,
2. For every $w \in W$, u is divisible by some prime $p \in J_w$.

To estimate the size of $|U|$ we require the following corollary of the combinatorial sieve of Rosser:

Proposition 1 *For every $0 < \beta < 1$, there exists $0 < \tau < 1$ so that for x sufficiently large, the following holds: Suppose P is a subset of the primes $\leq x^{\tau}$, and let Z be the set of all integers $\leq x$ not divisible by any prime $p \in P$. Then, if we let*

$$\Delta \ = \ \prod_{p \in P} \left(1 - \frac{1}{p} \right),$$

we will have

$$x(1-\beta)\Delta \;<\; |Z| \;<\; x(1+\beta)\Delta.$$

Now, to estimate the size of $|U|$, for each indexing set $V' \supseteq V$, we let $Z(V')$ denote the set of all integers $z \leq x$ not divisible by any prime $p \in J_v$, for any $v \in V'$. Then, by a simple inclusion-exclusion argument we have

$$|U| \;=\; \sum_{V' \supseteq V} (-1)^{|V'|-|V|} |Z(V')|.$$

In order to apply the proposition, we first estimate

$$\begin{aligned}
\Delta \;&=\; \prod_{v \in V'} \prod_{\substack{p \in J_v \\ p \text{ prime}}} \left(1 - \frac{1}{p}\right) \\
&=\; \prod_{v \in V'} \exp\left(- \sum_{\substack{p \in J_v \\ p \text{ prime}}} \frac{1}{p} \;+\; O\left(\frac{1}{x^{\gamma^j/2}}\right)\right) \\
&=\; \prod_{v \in V'} \exp\left(-\log 2 + O\left(\frac{1}{\gamma^j \log x}\right)\right) \\
&=\; \frac{1}{2^{|V'| + O(k/\gamma^j \log x)}}.
\end{aligned}$$

Note that this last expression is asymptotically $2^{-|V'|}$.

It now follows that from the proposition above, and from our equation for $|U|$ that for fixed ϵ, β, if γ is sufficiently small and x is sufficiently large, then

$$\begin{aligned}
|U| \;&=\; x\left(1 + O\left(\frac{1}{\log x}\right)\right) \sum_{V' \supseteq V} (-1)^{|V'|-|V|} \frac{1}{2^{|V'|}} \;+\; O\left(2^k \beta x\right) \\
&=\; \frac{x}{2^{|V|}} \left(1 + O\left(\frac{1}{\log x}\right)\right) \sum_{V'' \subseteq \{1,2,\ldots,k\} \setminus V} \frac{(-1)^{|V''|}}{2^{|V''|}} \;+\; O\left(2^k \beta x\right) \\
&=\; \frac{x}{2^k} \;+\; O\left(2^k \beta x\right).
\end{aligned}$$

The implied constant in this last big-oh depends on γ and ϵ.

So, the number of integers $s \leq x$ such that $h(s)$ is more than $k^{2/3}$ away from $k/2$ is

$$\sum_{\substack{V \subseteq \{1,2,\ldots,k\} \\ ||V|-k/2| \ > \ k^{2/3}}} \left(\frac{x}{2^k} + O\left(2^k \beta x\right) \right) \ = \ \frac{x}{2^k} \sum_{\substack{0 \leq j \leq k \\ |j-k/2| > k^{2/3}}} \binom{k}{j} \ + \ O\left(4^k \beta x\right).$$

Now, the sum over these binomial coefficients can be shown to be smaller than $2^k/k^2$; and so, for $\epsilon_0 < 1/5$, by choosing β and γ sufficiently small, we will have for x sufficiently large that there can be at most

$$\frac{x}{k^2} + O\left(4^k \beta x\right) \ < \ \frac{\epsilon x}{5}$$

integers $s \leq x$ with $|h(s) - k/2| > k^{2/3}$. This then proves that (13) holds for all but at most $\epsilon x/4$ elements of S.

To finish the proof of the theorem we will focus on the following function:

$$f(s) \ = \ \prod_{j=1}^{k} \left(\gamma^{2j} \log^2 x \ + \ \left(\sum_{\substack{p \in I_j, \ p|s \\ p \text{ prime}}} \log p \right)^2 \right).$$

First, let us calculate the sum of $f(s)$ over all $s \leq x$: This sum equals

$$\sum_{V \subseteq \{1,2,\ldots,k\}} (\log x)^{2|V|} \prod_{v \in V} \gamma^{2v} \sum_{s \leq x} \prod_{w \in \{1,2,\ldots,k\} \setminus V} \left(\sum_{\substack{p \in I_w, \ p|s \\ p \text{ prime}}} \log p \right)^2.$$

We now work with this inner sum by fixing a $V \subseteq \{1, 2, ..., k\}$, and letting $W = \{1, 2, ..., k\} \setminus V$, and then this inner sum is

$$\begin{aligned} &\sum_{s \leq x} \prod_{w \in W} \left(\sum_{\substack{p \in I_w, \ p|s \\ p \text{ prime}}} \log p \right)^2 \\ &\quad \leq \ x \prod_{w \in W} \left(2 \sum_{\substack{p_1 < p_2; \ p_1, p_2 \in I_w \\ p_1, p_2 \text{ prime}}} \frac{(\log p_1)(\log p_2)}{p_1 p_2} \ + \ \sum_{\substack{p \in I_w \\ p \text{ prime}}} \frac{\log^2 p}{p} \right) \end{aligned}$$

$$
\begin{aligned}
&= x \prod_{w \in W} \left(\left(\sum_{\substack{p \in I_w \\ p \text{ prime}}} \frac{\log p}{p} \right)^2 + \sum_{\substack{p \in I_w \\ p \text{ prime}}} \frac{\log^2 p}{p} \right) \\
&\le x \prod_{w \in W} \left(\left(\sum_{\substack{p \le x^{\gamma^w/2} \\ p \text{ prime}}} \frac{\log p}{p} \right)^2 + \sum_{\substack{p \le x^{\gamma^w/2} \\ p \text{ prime}}} \frac{\log^2 p}{p} \right) \\
&= x \prod_{w \in W} \left(\frac{\gamma^{2w}}{4} \log^2 x + \frac{\gamma^{2w}}{8} \log^2 x + O(\gamma^w \log x) \right) \\
&= x \left(1 + O\left(\frac{1}{\log x} \right) \right) \left(\frac{3 \log^2 x}{8} \right)^{|W|} \gamma^{2 \sum_{w \in W} w},
\end{aligned}
$$

where the implied constant in this last big-oh depends on γ and $|W| \le k$.

We then get that the sum of $f(s)$ over $s \le x$ is at most

$$
\begin{aligned}
&\gamma^{k(k+1)} x \left(1 + O\left(\frac{1}{\log x} \right) \right) (\log x)^{2k} \sum_{V \subseteq \{1,2,\ldots,k\}} \left(\frac{3}{8} \right)^{k-|V|} \\
&= \gamma^{k(k+1)} x \left(1 + O\left(\frac{1}{\log x} \right) \right) (\log x)^{2k} \left(\frac{11}{8} \right)^k .
\end{aligned}
$$

Before we bound $f(t)$ from below for an arbitrary $t \in T$ that fails to be divisible by any $n \in N$, we make a general observation: If $p \ge x^\delta$ is any prime divisor of such a t, then we must have that

$$
\frac{x}{\prod_{\substack{q \ge p,\ q | t \\ q \text{ prime}}} q} \ > \ p,
$$

for otherwise t is divisible by some integer $n \in N$. Now, since each member of $t \in T$ is at least $\epsilon x/100$, we deduce that

$$
\begin{aligned}
\prod_{\substack{q \le p,\ q^a || t \\ q \text{ prime}}} q^a \quad &= \quad \frac{t}{\prod_{\substack{q \ge p,\ q | t \\ q \text{ prime}}} q} \\
&\ge \quad \frac{\epsilon x}{100 \prod_{\substack{q \ge p,\ q | t \\ q \text{ prime}}} q} \\
&> \quad \frac{\epsilon p}{100}. \qquad (14)
\end{aligned}
$$

Now let j be one of the $> k/2 - k^{2/3}$ indices for which t is not divisible by any prime $p \in J_j$. Further, suppose that j is not the smallest index, which guarantees that t is divisible by at least some prime greater than x^{γ^j}. Then, the fact that (14) must hold implies

$$\prod_{\substack{q \le x^{\gamma^j/2},\ q^a \| t \\ q \text{ prime}}} q^a \;=\; \prod_{\substack{q \le x^{\gamma^j},\ q^a \| t \\ q \text{ prime}}} q^a \;>\; \frac{\epsilon x^{\gamma^j}}{100}.$$

Taking logs of both sides gives

$$\sum_{\substack{q \le x^{\gamma^j/2},\ q^a | t \\ q \text{ prime}}} \log q \;>\; \gamma^j \log x \;+ \log(\epsilon/100)$$

Now, from our assumption (11) we then deduce

$$\begin{aligned}
\sum_{\substack{q \in I_j,\ q|t \\ q \text{ prime}}} \log q \;&=\; \sum_{\substack{q \le x^{\gamma^j/2},\ q^a | t \\ q \text{ prime}}} \log q \;-\; \sum_{\substack{q \le x^{\gamma^{j+1}},\ q^a | t \\ q \text{ prime}}} \log q \\
&>\; \gamma^j (1 - 5\gamma k^2) \log x \;+\; \log(\epsilon/100) \\
&>\; \gamma^j (1 - \epsilon/2) \log x, \qquad (15)
\end{aligned}$$

for a fixed ϵ and γ sufficiently small. Thus, if we let X denote the set of indices j such that t is not divisible by any prime $p \in J_j$, then

$$\begin{aligned}
f(t) \;&\ge\; \prod_{j \in X} \left(\gamma^{2j} \log^2 x + (1-\epsilon/2)^2 \gamma^{2j} \log^2 x\right) \prod_{j \in \{1,2,\ldots,k\} \setminus X} \gamma^{2j} \log^2 x \\
&\ge\; \gamma^{k(k+1)} (\log^{2k} x)(2-\epsilon)^{k/2 - k^{2/3}}.
\end{aligned}$$

So, if we let Y be the integers $t \in T$ not divisible by any $n \in N$, then we have

$$\begin{aligned}
|Y| \gamma^{k(k+1)} (\log^{2k} x)(2-\epsilon)^{k/2-k^{2/3}} \;&<\; \sum_{y \in Y} f(y) \;<\; \sum_{s \le x} f(s) \\
&<\; x\gamma^{k(k+1)} (\log^{2k} x) \left(\frac{11}{8}\right)^k \\
&\qquad + O(x \log^{2k-1} x).
\end{aligned}$$

So, if ϵ_0 were sufficiently small (so as to make k sufficiently large, and ϵ sufficiently small), then we would clearly have $|Y| < \epsilon x/2$. Therefore, all but $\epsilon x/2$ integers $t \in T$ is divisible by some $n \in N$ once ϵ_0 is sufficiently small; and therefore, all but at most ϵx integers $\leq x$ are divisible by some $n \in N$. ■

References

[1] R. C. Baker and G. Harman, *Shifted Primes without Large Prime Factors* Acta Arith. **83** (1998), 331-361.

[2] A. Balog and A. Sárközy, *On sums of Integers Having Small Prime Factors. I, II*, Studia Sci. Math. Hungar. **19** (1984), 35-47.

[3] ——————————-, *On Sums of Sequences of Integers. I*, Acta Arith. **44** (1984), 73-86.

[4] ——————————-, *On Sums of Sequences of Integers. II*, Acta Math. Hungar. **44** (1984), 169-179.

[5] ——————————-, *On Sums of Sequences of Integers. III*, Acta Math. Hungar. **44** (1984), 339-349.

[6] R. de la Bretèche, *Sommes sans Grand Facteur Premier [Sums without Large Prime Factors]* Acta Arith. **88** (1999), 1-14.

[7] Dickman, *On the Frequency of Numbers Containing Prime Factors of a Certain Relative Magnitude* Ark. Mat. Astr. Fys **22** (1930), 1-14.

[8] P. Erdős, H. Maier, and A. Sárközy, *On the Distribution of the Number of Prime Factors of Sums* $a + b$ Trans. Amer. Math. Soc. **302** (1987), 269-280.

[9] J. B. Friedlander, *Shifted Primes without Large Prime Factors*, NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci., 265, Kluwer Acad. Publ., Dordrecht, 1989.

[10] C. Pomerance, A. Sárközy, and C. L. Stewart, *On Divisors of Sums of Integers. III.* Pacific J. Math. **133** (1988), 363-379.

[11] A. Sárkőzy and C. L. Stewart, *On Divisors of Sums of Integers. I* Acta Math. Hungar. **48** (1986), 147-154.

[12] ———————————, *On Divisors of Sums of Integers. II.*, J. Reine Angew. Math. **365** (1986), 171-191.

[13] A. Sárkőzy and C. L. Stewart, *On Divisors of Sums of Integers. IV.* Canad. J. Math. **40** (1988), 788-816.

[14] ———————————, *On Divisors of Sums of Integers. V.* Pacific J. Math. **166** (1994), 373-384.